18th ICCRTS

C2 in Underdeveloped, Degraded and Denied Operational Environments

Title: A Visual Analytic for Improving Human Terrain Understanding

Topic Area 3: Data, Information, and Knowledge
Topic Area 4: Collaboration, Shared Awareness, and Decision Making

Authors: Dr. Timothy Hanratty,
Mr. Eric Heilman,
Mr. John Richardson,
Mr. Mark Mittrick
US Army Research Laboratory
RDRL-CII-C
Building 321
Aberdeen Proving Ground, MD 21005

POC: Dr. Timothy Hanratty
timothy.hanratty@us.army.mil
Phone: 410-278-3084

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **JUN 2013**	2. REPORT TYPE	3. DATES COVERED **00-00-2013 to 00-00-2013**

4. TITLE AND SUBTITLE **A Visual Analytic for Improving Human Terrain Understanding**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Army Research Laboratory,Tactical Information Fusion Branch,Computational Information Science Directorate,Aberdeen Proving Ground,MD,21005**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the 18th International Command & Control Research & Technology Symposium (ICCRTS) held 19-21 June, 2013 in Alexandria, VA.

14. ABSTRACT
Visual analytics are a growing area of research within the Department of Defense that targets the effective interleaving of analytical reasoning with interactive interfaces. Capitalizing on the human capacity for spatial reasoning, visual analytics enhance the decision maker?s understanding of the underlying decision space by augmenting the assimilation of complex relationships. Towards this end, the US Army Research Laboratory has developed a software application that complements traditional social network analysis and allows for improved understanding of the human terrain. The software application, called the Heterogeneous Data Proximity Tool (HDPT), combines a data dimensionality reduction routine with multidimensional scaling to produce an analytic to visually interrogate the similarity or dissimilarity among a given set of high valued individuals. The HDPT starts with a reference data set representing individuals with known group affiliations: insurgents, innocents, and criminals. As intelligence is collected about new individuals within an area of operation, HDPT computes and plots their relative positions with regard to the existing reference data. The resulting analytic portrays the relative position of the new individual?s organizational orientation within the known human terrain. This paper presents the development of HDPT and the results from its participation in the FY12 Product Director Command, Control, Communications, Computers, Intelligence, Surveillance, Reconnaissance (C4ISR) & Network Modernization’s Event 2012 (E12) field exercise.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**29**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

A Visual Analytic for Improving Human Terrain Understanding

Timothy Hanratty, Eric Heilman, Mark Mittrick and John Richardson
Tactical Information Fusion Branch
Computational Information Science Directorate
US Army Research Laboratory
Aberdeen Proving Ground, MD, USA
timothy.hanratty@us.army.mil

Abstract — Visual analytics are a growing area of research within the Department of Defense that targets the effective interleaving of analytical reasoning with interactive interfaces. Capitalizing on the human capacity for spatial reasoning, visual analytics enhance the decision maker's understanding of the underlying decision space by augmenting the assimilation of complex relationships. Towards this end, the US Army Research Laboratory has developed a software application that complements traditional social network analysis and allows for improved understanding of the human terrain. The software application, called the Heterogeneous Data Proximity Tool (HDPT), combines a data dimensionality reduction routine with multidimensional scaling to produce an analytic to visually interrogate the similarity or dissimilarity among a given set of high valued individuals. The HDPT starts with a reference data set representing individuals with known group affiliations: insurgents, innocents, and criminals. As intelligence is collected about new individuals within an area of operation, HDPT computes and plots their relative positions with regard to the existing reference data. The resulting analytic portrays the relative position of the new individual's organizational orientation within the known human terrain. This paper presents the development of HDPT and the results from its participation in the FY12 Product Director Command, Control, Communications, Computers, Intelligence, Surveillance, Reconnaissance (C4ISR) & Network Modernization's Event 2012 (E12) field exercise.

Keyword: Visual Analytic, Multidimensional Scaling, Dimensional Reduction, Situation Awareness.

I. INTRODUCTION

The US Army Research Laboratory's (ARL) Tactical Information Fusion Branch has as its mission the research and development of advanced information analytics to assist Soldiers in determining, utilizing and sharing relevant information and improving the synthesis of data to decisions. Challenging these efforts are not only the unprecedented increase in the types and amount of information available, but the human judgment necessary to evaluate it in the presence of incomplete and inconsistent information within time-critical environments [1,2]. Required are innovated methods that allow the efficient and effective transformation of data to information to knowledge. One process proffered to tackle this challenge is visual analytics.

Visual analytics is a growing area of research within the Department of Defense that targets the effective interleaving of analytical reasoning with interactive interfaces. Capitalizing on the human capacity for spatial reasoning, visual analytics enhance the decision maker's understanding of the underlying decision space by augmenting the assimilation of complex relationships [3,4]. Visual analytics is a multidisciplinary field that has yield significant results in an array of paradigms to include business, medicine, and defense [1, 5].

Towards this end and the subject of this report, the Tactical Information Fusion Branch has developed a visual analytic application, entitled the Heterogeneous Data-reduction Proximity Tool (HDPT). The goal of the HDPT is to complement traditional social network analysis by allowing the exploitation of a social network based upon the calculated similarity of individuals (against a known reference set) as opposed to the traditional organizational structure. The result is an improved understanding of the human terrain that effectively incorporates qualitative and quantitative information into the decision making process.

This report documents the development and evaluation of the HDPT visual analytic application. In Section 2, background information on the statistical approach undertaken and the rational for choosing the particular proximity calculation is provided. In section 3, the system-level design and instantiation of the HDPT as a web application linked to the Distributed Common Ground Systems - Army (DSGS-A) is discussed. The results of from HDPT utilization in an Army field study are presented in section 4. The paper concludes with lessons learned and the way forward in Section 5.

II. BACKGROUND

For many military applications, extracting knowledge from high-dimensional data sets is a persistent and complicated task. This is especially true when the data sets are of mixed-data type, wherein the attributes defining the objects to be compared take on values from differing measurement scales. Moreover, the data of interest are typically amorphous; i.e. not linked to an explicit theory to assist the researcher in making inferences or predicting structure. For these reasons, HDPT uses multidimensional scaling (MDS) for visualizing the structure of the data and Gower's similarity coefficient as the algorithm for calculating the proximity matrices. The following sections provide a brief background on both MDS and Gower.

A. Multidimesional Scaling

Originating out of the fields of mathematical psychology and social sciences, multidimensional scaling (MDS) is a data analysis approach used to visually interrogate the similarity or dissimilarity between the pair-wise "distances" among a given set of objects [6,7,8,9,10,11]. The values of the distances, sometimes called proximity measures or similarity measures, can be obtained either as perceived subjective measures or calculated objectively within the pair-wise comparison of the given set of objects. Most often, the objects are vectors of the form $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$ with the components $\mathbf{x}_k$ known as attributes, variables, or factors collectively providing the basis for comparison of the objects. Given a similarity matrix for a set of objects, each object is projected as a point in n-space – arranged so the distances between the objects have the strongest possible relation to the similarity matrix. The intrinsic power of MDS is that it allows the dimensionality reduction of a complex n-space to a human interpretable 2– or 3–dimensional space. It is this projection that promotes the exploratory analysis of data's hidden structure.

The following example from Forrest Young's *Understanding Multidimensional Scaling* [7] uses the mileage between 10 American cities, shown in Table 1, as the objective similarity measure.

Atlanta	Chicago	Denver	Houston	Los Angeles	Miami	New York	San Francisco	Seattle	Washington, DC	
0	587	1212	701	1936	604	748	2139	2182	543	Atlanta
587	0	920	940	1745	1188	713	1858	1737	597	Chicago
1212	920	0	879	831	1726	1631	949	1021	1494	Denver
701	940	879	0	1374	968	1420	1645	1891	1220	Houston
1936	1745	831	1374	0	2339	2451	347	959	2300	Los Angeles
604	1188	1726	968	2339	0	1092	2594	2734	923	Miami
748	713	1631	1420	2451	1092	0	2571	2408	205	New York
2139	1858	949	1645	347	2594	2571	0	678	2442	San Francisco
2182	1737	1021	1891	959	2734	2408	678	0	2329	Seattle
543	597	1494	1220	2300	923	205	2442	2329	0	Washington, DC

Table 1: Between-City Mileage Similarity Matirix [7]

The associated MDS 2-D visualization output would appear something like that found in figure 1. Note: the geometric model allows one to discern the underlying structure and allow human interpretation.

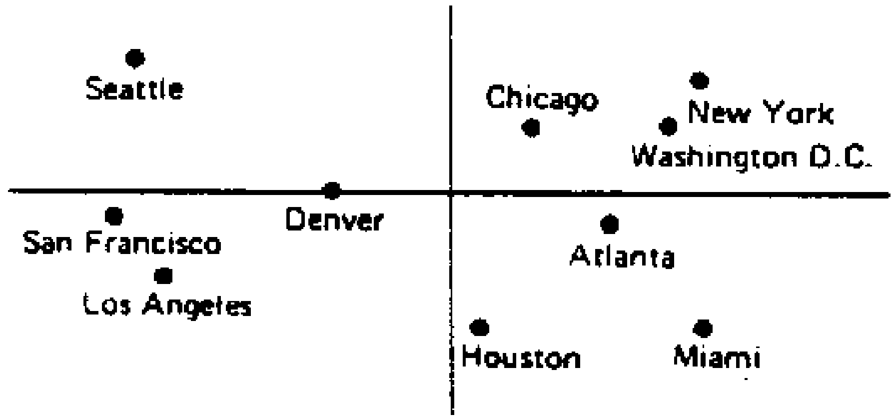


Figure 1: MDS US City Example [7]

Differing from other forms of multivariate statistics, specifically principal component analysis (PCA), MDS does not constrain the data to be normally distributed. With that understanding, it becomes apparent that the hidden power behind meaningful MDS analysis is found in the construction of the similarity matrix and its projection into the reduced space.

B. Gower Similarity Coefficient

For ideally calculating a similarity matrix, all of the defining attributes should be of the same data type [12]. Unfortunately, for many real-world problems disparate scales of measure are common place making matrix calculation problematic. One of first to confront the combination of quantitative and qualitative (mixed scales of measure) was John Gower [13]. Given an array of objects with k attributes the global similarity value ($\mathbf{S}_{ij}$) between two objects is defined as the summation of the individual attribute similarities ($\mathbf{s}_{ijk}$) multiplied by a possible weighting factor. Here, $\mathbf{s}_{ijk}$ corresponds to the measure of local similarity assigned to the object pair ($\mathbf{X}_i$, $\mathbf{X}_j$), restricted to attribute k. The summation of the individual similarities is divided by the summation across all weights. Gower's similarity coefficient equation shown below, allows for the weighing of individual attributes and the possibility of missing data.

$$S_{ij} = \frac{\sum_{1}^{k} s_{ijk}\, w_k}{\sum_{1}^{k} w_k}\text{; } w_k \text{ weight assigned to an individual attributes}$$

The classic calculation for individual similarities is given below - where $\mathbf{X}_{ik}$ and $\mathbf{X}_{jk}$ are the $\boldsymbol{K}^{th}$ attribute for objects $\mathbf{X}_i$ and $\mathbf{X}_j$ respectively. $\boldsymbol{R}_k$ is defined as the range for that particular quantitative attribute. In recent years, numerous extensions to similarity measurement calculations have been attempted in wide array of subject areas from image processing to medical informatics. Approaches taken include but are not limited to rough sets [14], fuzzy logic [15], and ordinal extensions [16].

$$s_{ijk} = \begin{cases} \begin{cases} 1, & \text{if } x_{ik} = x_{jk} \\ 0, & \text{if } x_{ik} \neq x_{jk} \end{cases} & \text{,k is qualitative} \\ 1 - \dfrac{|x_{ik} - x_{jk}|}{R_k} & \text{,k is quantitative} \end{cases}$$

The following section details the design and development of the HDPT visual analytic application. Specifically, the HDPT's development for concept demonstration in the US Army's Command, Control, Communication, Computers, Intelligence, Surveillance and Reconnaissance (C4ISR) On-The- Move (OTM) 2012 exercise (E12) and associated scenario.

III. HDPT SYSTEM OVERVIEW

While there exists research into the development of qualitative and quantitative similarity analysis, few have been effectively coupled with a visualization framework and fewer still have been interactively coupled with a tactical military decision support tool. Toward that end, HDPT is the software instantiation of a visual analytic technique that effectively combines MDS with the flexibility of the mixed-scale Gower similarity calculation. The targeted area of interest for this instantiation is the assessment of individuals within a tactical social network. This specific implementation was designed as a concept demonstration for the Product Director Command, Control, Communications, Computers, Intelligence, Surveillance, Reconnaissance (C4ISR) & Network Modernization's Event 2012 (E12) exercise and was integrated with the US Army's Distributed Common Ground System – Army (DCGS-A) program as data feed.

At a high level of abstraction, the concept of operation for this exercise was as follows. The HDPT started with a reference data set that represented the characteristics of individuals with known group affiliations: insurgents, innocents, and criminals. As intelligence was collected about new individuals within an area of operation, HDPT computed their similarity to the reference data set and plotted their relative positions in an associated 3-D visualization space. The resulting analytic portrayed the relative position of the new individual's orientation within the known human terrain (insurgent vs innocent vs criminal). The goal of the tool is to give a military analyst improved understanding of the local human environment and assist in defining future information requests.

Shown in Figure 2 are three major components that make up the HDPT system:

- *the HDPT Web Application,*
- *the DCGS-A Global Graph, and*
- *the statistics engine.*

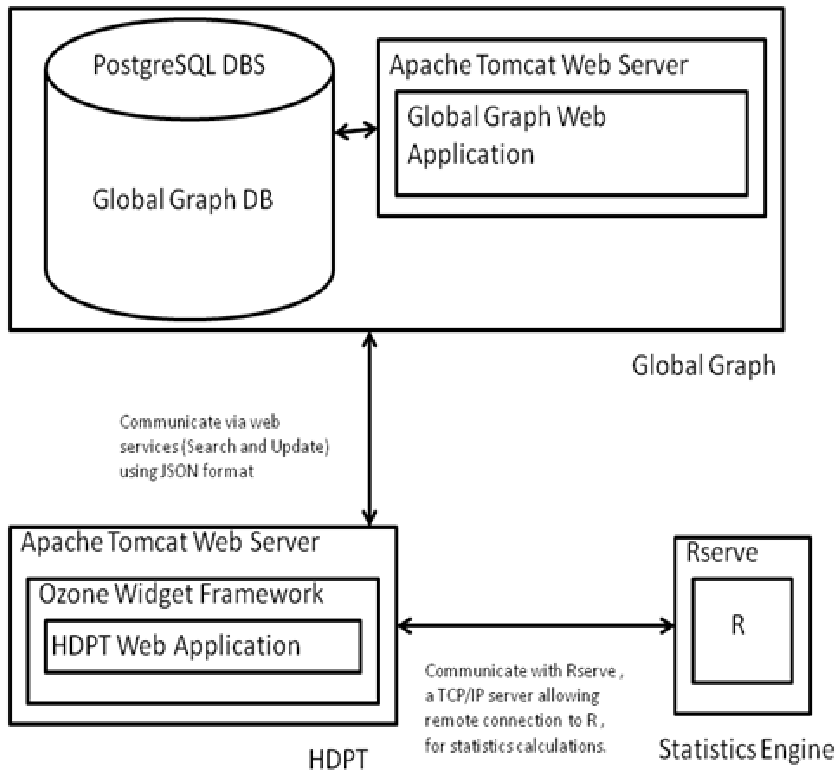


Figure 2: HDPT System Diagram

HDPT is a web application that is deployable within the Ozone Widget Framework (OWF). It accesses the DCGS-A Global Graph via web service as its data source. To calculate similarity it uses the Gower Similarity and Multidimensional Scaling algorithms contained in the R statistical computing environment. The Global Graph and OWF are both products by Potomac Fusion, Inc., and part of the DCGS-A program of record (POR). The principal HDPT component and subject of the remainder of this paper is the *HDPT Web Application.* The following subsections outline the design of the *HDPT Web Application* as demonstrated at the C4ISR E12exercise.

A. HDPT Web Application

The *HDPT web application* is the core of the HDPT system and serves as the user interface for performing similarity analysis. The primary components of *the HDPT web application* used in E12 are displayed in Figure 3 and include the following: (1) HDPT Menu Bar, (2) Search Window Panel, and (3) Plot Window Panel.

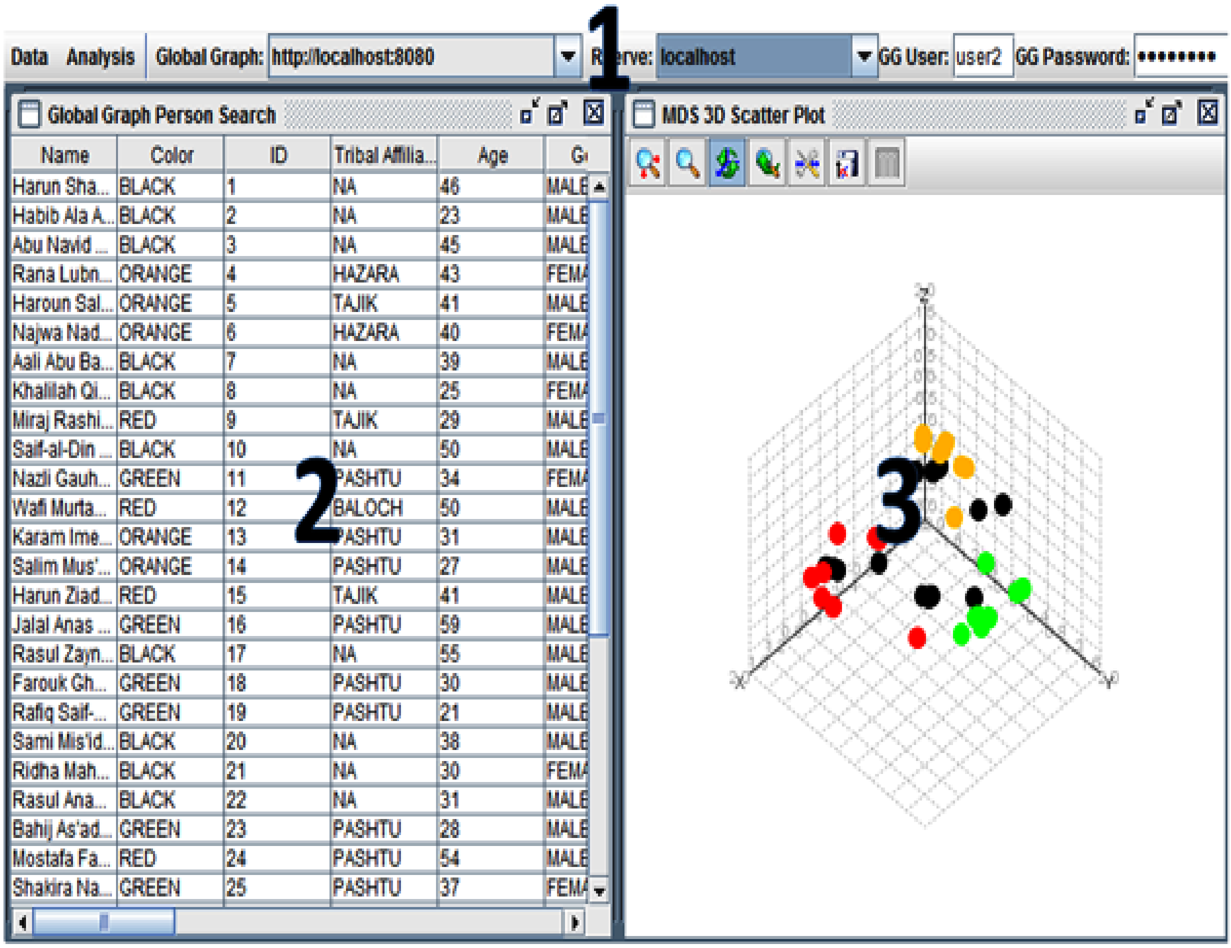


Figure 3: HDPT Web Application Components: (1) Menu Bar (2) Search Window (3) Plot Window

1) HDPT Menu Bar

The HDPT Menu Bar contains drop-down selections for loading data, creating the visual analytic and configuring the HDPT. The menu bar allows access to the *Data Menu*, *Analysis Menu*, and the *configuration* components.

- The *Data Menu* contains drop-down selections for loading data into the tool and propagating updates back to the data source. The data source used by HDPT during the E12 exercise was the DCGS-A Global Graph. In this exercise, an SQL version of the Global Graph was used that consisted of a PostgreSQL database and associated web services for searching and updating the database. A REST (Representational State Transfer) web service protocol was used for communication between HDPT and the data source via a JSON (JavaScript Object Notation) data structure. REST is a lightweight alternative protocol to mechanisms like SOAP, RPC or COBRA. Typically with REST, simple HTTP is used to make the connections. Likewise, JSON is a lightweight data-interchange format designed for exchanging human-readable structured text. The selection of these protocols greatly facilitated connection and interaction with the DCGS-A framework.

- The *Analysis menu* contains drop-down selections for plotting the visual analytic as an interactive 3D scatter plot and setting user preferences. The **Plot** menu action becomes available once there is an active search window panel. In addition to plotting, the *analysis menu* allows two user **preferences** to be modified: links displayed and attribute threshold. Both of these preferences change the way links between nodes in the plot are handled. By default, when interacting with the plot the user can right click on a node and links will be drawn to the three most similar reference nodes; the number of links drawn can be increased or decreased.

- The *configuration menu*, allows the user to specify server parameters for the Global Graph and Rserve. There is a dropdown menu for each server selection that contains many commonly used servers for the OTM exercise. In addition, mandatory fields are provided for users to enter their username and password for Global Graph access. HDPT uses these values for connecting to both the Global Graph REST web services and the R statistics engine.

2) Search Window Panel

The HDPT search window panel displays the dataset returned from a Global Graph search, as seen in Figure 4.

Global Graph Person Search

Name	Color ▲	ID	Tribal Affiliat...	Age	Gender	Marital Status
Aali Abu Bakr Karim	BLACK	11	NA	39	MALE	NA
Abu Navid Sultan	BLACK	32	PASHTU	42	MALE	NA
Amirah Sani El-Amin	BLACK	36	NA	48	FEMALE	NA
Habib Ala Ahmed	BLACK	34	NA	23	MALE	NA
Harun Shahzad El-Mofty	BLACK	9	NA	46	MALE	NA
Hussain Mansoor El-Hashem	BLACK	22	NA	32	MALE	NA
Ikram I'timad Abdullah	BLACK	16	HAZARA	20	MALE	MARRIED
Khalilah Qismat Amirmoez	BLACK	13	NA	25	FEMALE	NA
Rasul Anass Zaman	BLACK	4	NA	31	MALE	NA
Rasul Zayn Mohammed	BLACK	20	NA	55	MALE	NA
Ridha Mahdi El-Mofty	BLACK	19	NA	30	FEMALE	NA
Saif-al-Din Jinan Hakim	BLACK	24	NA	50	MALE	NA
Sami Mis'id El-Ghazzawy	BLACK	5	NA	38	MALE	NA
Yusuf Mehmud Samara	BLACK	15	NA	42	MALE	NA
Zaman Noor Hakim	BLACK	7	NA	34	MALE	NA
Bahij As'ad Tawfeek	GREEN	6	PASHTU	28	MALE	MARRIED
Farouk Ghayth El-Ghazzawy	GREEN	26	PASHTU	30	MALE	MARRIED
Gabr Hussein Ahmed	GREEN	1	BALOCH	35	MALE	SINGLE
Hashim Fouad Admad	GREEN	3	TAJIK	35	MALE	SINGLE
Jalal Anas Kader	GREEN	18	PASHTU	59	MALE	SINGLE
Nazli Gauhar Ajam	GREEN	8	PASHTU	34	FEMALE	MARRIED
Rafiq Saif-al-Din Karim	GREEN	33	PASHTU	21	MALE	SINGLE
Shakira Nashwa Abujamal	GREEN	37	PASHTU	37	FEMALE	MARRIED
Haroun Salih Abdullah	ORANGE	39	TAJIK	41	MALE	SINGLE
Jinan Qadir El-Ghazzawy	ORANGE	27	BALOCH	44	MALE	SINGLE
Karam Imen Boulos	ORANGE	29	PASHTU	31	MALE	SINGLE
Majid Rusul Abujamal	ORANGE	25	TAJIK	38	MALE	SINGLE
Mostafa Tufayl Karimi	ORANGE	28	PASHTU	30	MALE	SINGLE
Najwa Nadia Saqqaf	ORANGE	17	HAZARA	40	FEMALE	SINGLE
Rana Lubna Samara	ORANGE	12	HAZARA	43	FEMALE	SINGLE
Salim Mus'ad Hakim	ORANGE	2	PASHTU	27	MALE	SINGLE
Adam Abdur-Rashid Abdullah	RED	10	BALOCH	35	MALE	MARRIED
Harun Ziad Boulos	RED	14	TAJIK	41	MALE	MARRIED
Miraj Rashid Karimi	RED	21	TAJIK	29	MALE	MARRIED
Mostafa Farooq Darzi	RED	38	PASHTU	54	MALE	MARRIED
Nasir Baki Saab	RED	23	HAZARA	34	MALE	SINGLE
Sulaiman Badr Muhammad	RED	31	PASHTU	44	FEMALE	SINGLE
Wafi Murtada Hakim	RED	30	BALOCH	50	MALE	MARRIED
Ziyad Guda Sultan	RED	35	HAZARA	34	MALE	MARRIED

Figure 4: Search Window Panel

The window consists of a table where each row is a person and each column is an attribute. The result returned from the Global Graph web search service is a JSON structure containing all matching people. HDPT processes the JSON structure into a separate Java object for each person. The Search window has a limited number of user interactive features. In addition to attribute editing, which was discussed previously as part of the **Analysis Menu** actions, the Search Window panel allows rows to be sorted according to the values in any column. Clicking the mouse on the column heading will cause the rows to sort alphabetically (words) or number

order (digits) according to the data in that column. For example, the search panel shown in Figure 4 is sorted by the “Color” column. Finally, if the plot window is currently active, clicking on a row in the search window will highlight the node that corresponds to that row in the 3D scatter plot; changing that node’s color to a neutral yellow.

3) Plot Window Panel

The Plot Window Panel displays a 3D scatter plot of MDS results. The development of the visualization used the JMathPlot open source graphics library. A sample Plot Window Panel from the E12 exercise is shown in figure 5. In this case, the node under investigation is high-lighted using a neutral yellow color and has the three ‘most’ similar reference nodes linked with straight lines; sharing similarity between two nodes from the criminal set (orange color) and one from the friendly set (green).

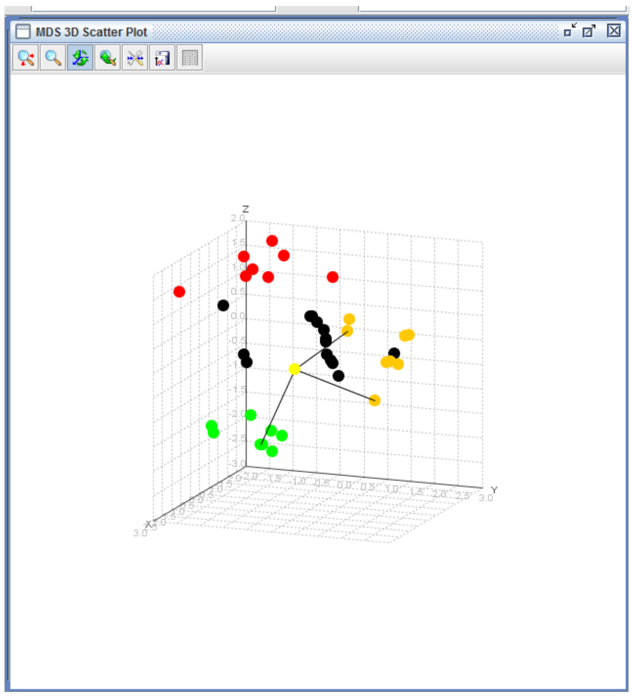


Figure 5: Plot Window Panel showing Similarity Links

B. HDPT Plot Window Panel Utility

In the E12 exercise the nodes were color coded with the following population schema: BLACK=Unknown, RED=Insurgent, GREEN=Friendly, and ORANGE=Criminal. As the exercise transpired data describing Unknown persons was updated in the Global Graph, either from the HDPT console or a mobile device ARL developed for soldiers’ use during field encounters. As the data associated with each BLACK node changed, the HDPT visual analytic also changed. Soldiers participating in the exercise as Intelligence Analysts were tasked with using the scatter plot to determine the disposition of BLACK nodes. That is, did projection of the dimensionally reduced characteristics of each BLACK node more likely belong to the RED, GREEN, or ORANGE population as determined by proximity (similarity).

To assist the Soldier’s understanding of the underlying decision space, HDPT provided a couple of important capabilities to the 3-D visual analytic that were used extensively throughout the E12

exercise. First, as shown in figures 6(a) and 6 (b), HDPT provided the ability to freely rotate the decision space along any axis. Projection of a 3-D decision space onto a 2-D screen can be problematic; objects that appear close to one another in 2-D can actually be far apart. The ability to rotate along any axis was critical to correctly interpreting the relation projections of the decision space.

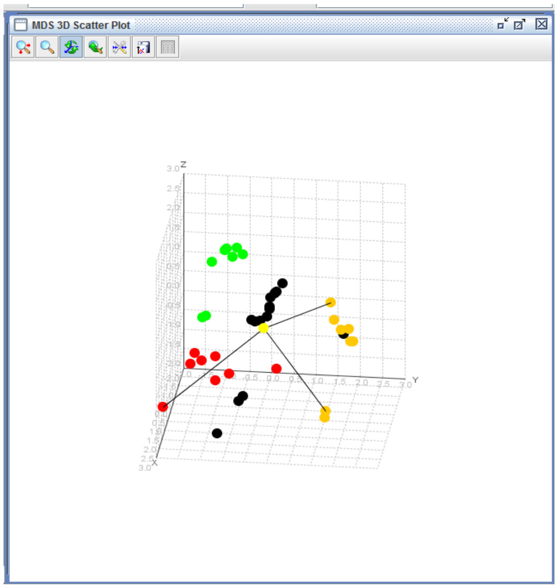


Figure 6(a) View before rotation

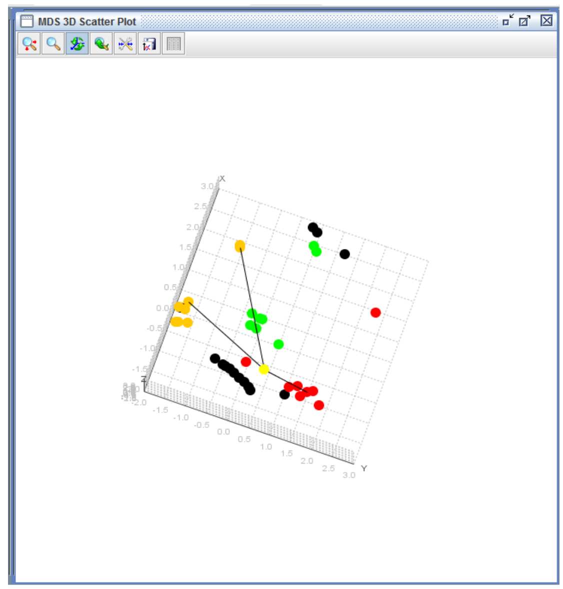


Figure 6(b) HDPT Rotated View

A second capability that was utilized at length was the zoom. As shown in figure 7(a) and 7(b), the zoom capability permitted users the ability to examine in finer detail the related nodes that were clustering close to the node in question.

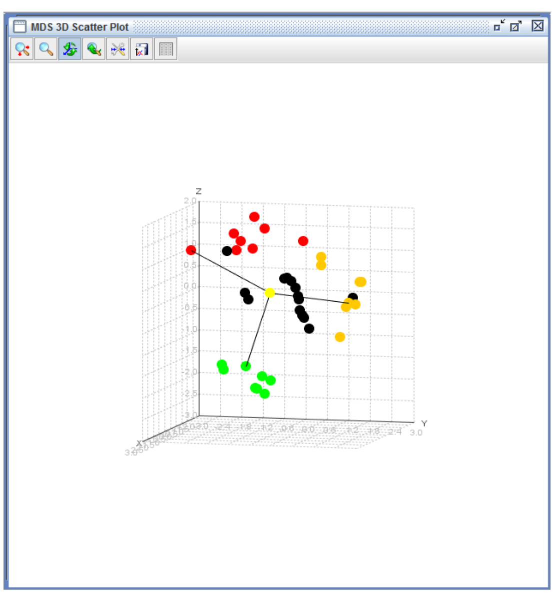


Figure 7(a) View before zoom

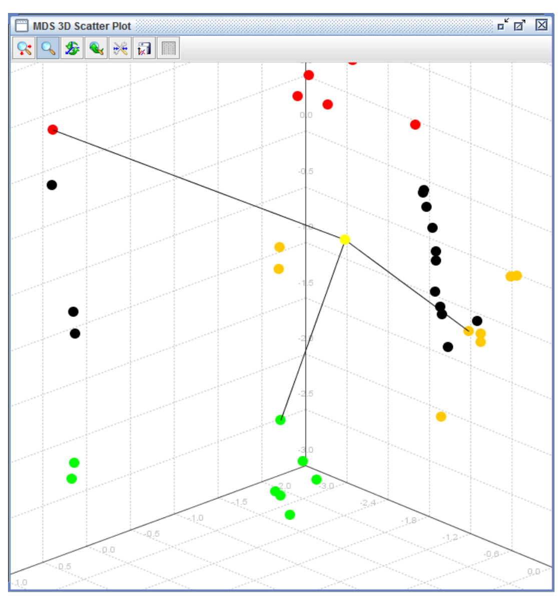


Figure 7(b) HDPT zoomed View

In this example, the node in question (yellow node) shows a high similarity to (3) neighboring nodes; one from each of the categorical types.

The Plot Window Panel also possessed several interactive features to assist in exploratory analysis. Left clicking on a node in the display both highlighted it and the corresponding row in the search window so that the node is referenced back to its original data. Second, as discussed previously, right clicking on a node will draw links from itself to 3 or more of its most similar reference node neighbors.

IV. HDPT C4ISR -E12 CONCEPT EVALUATION

The goal of the C4ISR E12 event is to provide stakeholders from across the Department of Defense assess to next-generation technologies. The annual event, held at Fort Dix, New Jersey, offers researchers a military relevant venue to assess, evaluate and validate emerging technologies and facilitate technology maturation and transition to the acquisition process. The following section details HDPT participation at the E12 event.

A. Reference Data Set Description

Table 2: HDPT Data Attributes

Tribal Affiliation	Education Level
Marital Status	Employment Type
Nationality	Military Record
Place of Birth	(person religion) Member of
EquipmentID	Skill
VehicleID	AddressID
Criminal Record	

The HDPT team created a multiple-dimensioned attribute data set describing persons affiliated with several groups presumed to be operational in the E12 exercise area. The groups included personnel considered to be non-hostile or friendly, overtly hostile (insurgent terrorists), and locally hostile (criminal). The attribute set, shown in Table 2, represents information collectable through combat questioning of individuals met during field encounters. To support HDPT analysis, each attribute can take the form of one or more specific values. For instance, the Martial Status attribute has a value of either Married or Not Married to represent the current condition of an individual. A full set of specific values for a person's attributes constitutes a unique data set representing that specific individual. For illustrative purposes, an example of a person data element set is shown in Figure 8. The subject's name is Bahij As'ad Tawfeek, a male subject. He is 28 years old, married and a member of the Pastun tribe. Bahji was born in the area and is an afghan national (HN is Host Nation). Soldiers have observed him wearing a military-like uniform and have associated him with a burgundy luxury sedan. He has no criminal record, is highly educated, and is considered to be holding down a white collar job as a mechanic serving in the local military. He belongs to a milder religious sect where he lives in Time Square Village.

Name	SubID	Location	Org	Age	Gen	TA	MS	Nat	POB	Equip	VehicleID	CR	Ed	Emp	MR	Rel	Sk	Add
Bahij As'ad Tawfeek		39.98 / -74.43	Friendly	28	M	Pastun	M	HN	BIA	Uniform	Burgundy Luxury Sedan	No	High	WC	SS	Mld	ME	TSV

Figure 8: Example of a person data element set

For the purposes of testing HDPT, the ARL team created thirty-nine attribute sets, similar to Figure 8. These were broken down into three reference groups each containing eight individuals representing the friendly, insurgent, and criminal organizations and fifteen sets representing initially unknown individuals. The reference groups formed node clusters in the HDPT visual plot, with each node representing one person. Figure 9 shows the HDPT display with both the insurgent (red nodes) and the friendly (green nodes) organizations circled. The black nodes represent unknown individuals. For ground truth purposes, the full set of attributes for each of the fifteen unknown persons placed their representative black nodes within an HDPT reference cluster. The removal of data from the thirteen fields listed in Table 2 causes the neutral alignment of black nodes for the initial HDPT display shown in Figure 9.

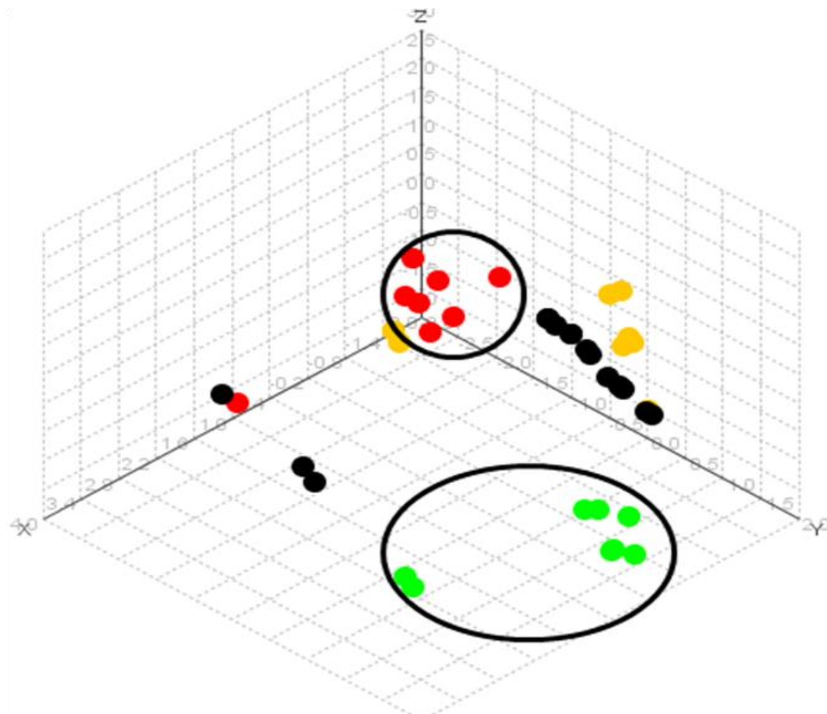


Figure 9: HDPT Reference Group Visual Display.

B. Scenario Inputs

During each day of the field exercise, Soldiers conducted a checkpoint and a presence patrol. It was during these missions that combat questioning of up to four threat actors (unknown individuals) was performed. The questioning was used to obtain data to populate the character's thirteen missing fields. Early in the exercise, the ARL team found that the realistic conditions of combat questioning most often yielded only a part of the data necessary for full analysis. To support a thorough study of HDPT capabilities, the ARL team injected data for eight unused characters to ensure that an optimal set of attributes was available for analysis by soldiers performing the military intelligence functions within the tactical operations center. Four of the characters were represented using the complete set of thirteen data attributes while the other four characters had data for only six attributes. The resulting data mix provided an ad hoc condition from the field data as well as a controlled condition supportive of HDPT ground truth analysis.

Data injects were divided into two sections: one for the morning mission and one for the afternoon mission. Figure 10 shows an example visual analytic of a daily progression of data found for a single character. The character node is colored yellow and has the three most similar nodes linked with straight lines. The left visual analytic represents data discovered in the morning of an exercise day while the visual analytic on the right is an accumulation of the morning data and data discovered during the afternoon. The progression of data discovery is representative of intelligence accumulation on persons of interest over a period of time. In this example, the morning data causes the character to appear friendly, but additional data from afternoon collection casts the character more as a criminal. The ground truth puts this character in the criminal organizational group.

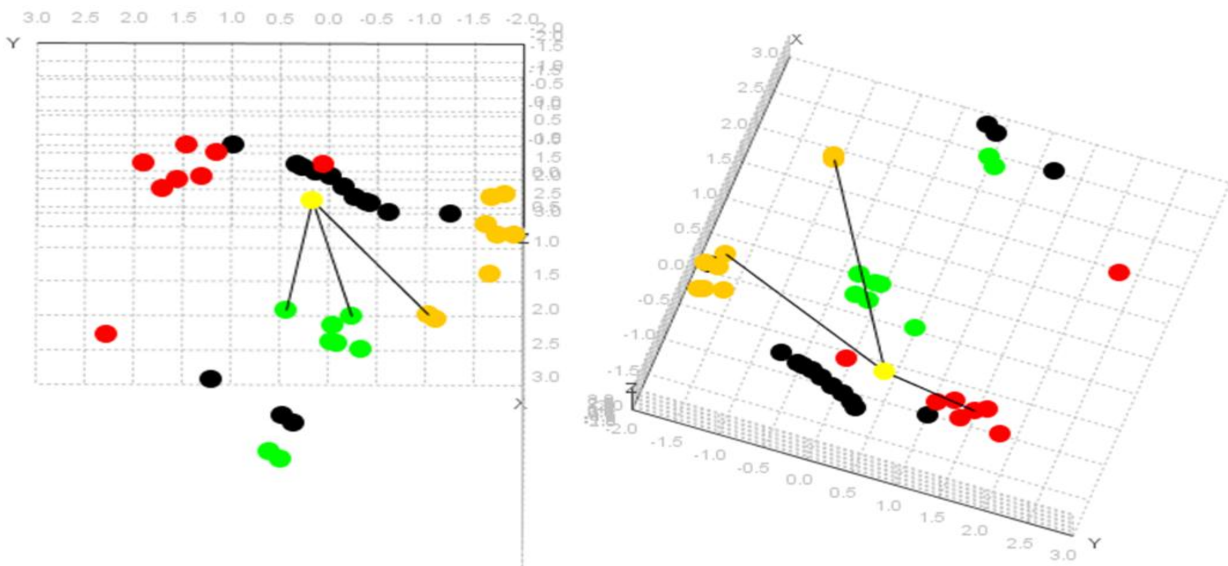


Figure 10: Progressive data discovery for a single unaligned person

C. Soldier Survey

To track the progression of the soldiers' understanding throughout an exercise day a survey form (see appendix A) was used. Using the HDPT visual analytic, each Soldier was asked to give an initial, mid-morning and mid-afternoon assignment for each of the unknown (black nodes) individuals association to a group and to weight that assignment on a scale from 1 to 5. The Soldiers' assignment as compared to character ground truth was used as a measure of HDPT's utility. Soldiers were also given the chance to comment on their understanding of character situational awareness based on the HDPT visual analytic, the utility of the tool and open comments for improvements.

D. Results

During the exercise, there were 11 different soldiers acting as intelligence analysts using the HDPT. Most of these soldiers, 10 out of the 11, were infantryman; typical of those found on a Company Intelligence Support Team. Few, 2 out of the 11, had military intelligence training; while most felt confident in their computer use abilities. Using a likert scale of 1 to 5 (where 1

equated to *very poor* and 5 equaled *very well*) the survey of the Soldiers' revealed the following concerning the use of the HDPT, the effectiveness of HDPT to track the evolving High Value Individual (HVI) intelligence picture and their opinion of the usefulness of HDPT in a tactical deployment. Most found that HDPT was easy to use; scoring usability a 4.18 out of a possible 5. The soldiers also thought that HDPT addresses a tactically useful function; scoring a 4.09 out of a possible 5. Some soldiers were concerned about HDPT's ability to provide an improved understanding of civilian personnel within the exercise area of operations; scoring a 3.8 or a possible 5. Concerns raised during the E12 event were reflected in the comments section of the survey.

When commenting on the HDPT, soldiers suggested the development of several features that were implemented during the exercise, these included: connection of the three persons closest to an unknown node under examination with straight line upon querying and an improved resolution of nodes that overlap in the HDPT visual analytic. Additionally, soldiers suggested several improvements are being developed currently, to include: highlighting of the HDPT person data spreadsheet corresponding to both a chosen node and the three closest nodes to that node and visual analytic rotation while in a zoomed portion of the visual analytic. As these suggestions originate from the ultimate field user of the HDPT, each will strengthen the relevance of the final product. The value of such interactions early within the technology creation process is crucial to ensure the creation of the highest quality tools for our soldiers' use.

S #	Age	MOS	Rank	Position	Yrs in	Color Blind	Compuer Use	# of Deploy	Intell Training	How Well used	Unk Person Sit Aware	HDPT Use
1	24	19K30	SGT	Tank Cdr	5	no	2 of 5	2	None	4 of 5	4 of 5	4 of 5
2	21	11B	PV2	SAW	0.75	no	5 of 5	0	None	5 of 5	3 of 5	3 of 5
3	26	11B	SGT	Sqd Ldr	5	no	4 of 5	2	None	4 of 5	3 of 5	3 of 5
4	24	11B	SPC	Team Ldr	5	no	4 of 5	1	None	4 of 5	4 of 5	3 of 5
5	19	11B	PFC	Infantryman	1	no	3 of 5	0	None	4 of 5	4 of 5	5 of 5
6	41	11B	SGT	Sqd Ldr	18.5	no	5 of 5	7	Yes	4 of 5	5 of 5	4 of 5
7	52	42A	LTC	S3	32	no	3 of 5	2	Yes	5 of 5	-	5 of 5
8	21	11B	PV2	Infantryman	1	no	4 of 5	0	None	5 of 5	4 of 5	4 of 5
9	22	11B	PFC	Infantryman	1	no	4 of 5	0	None	3 of 5	4 of 5	4 of 5
10	19	11**B**	PFC	Infantryman	1	no	4 of 5	0	None	4 of 5	4 of 5	5 of 5
11	20	11B	PV2	SAW	0.75	no	3 of 5	0	None	4 of 5	3 of 5	5 of 5
					Averages=		3.72			4.18	3.8	4.09

Table 3: HDPT Survey Result Table

Soldiers using the HDPT examined the similarity of each unknown node with two goals: first, to predict in which reference group an unknown node belonged, and second, to provide a confidence on that prediction with strength of from 1 (least confident) to 5 (most confident). There were two different soldiers using HDPT during each day of the exercise. As shown in Figure 11, the Soldiers' analysis of nodes with both full and partial attribute data resulted in a high percentage of correct predictions. In fact, as determined the strength coefficient, most analysis became progressively more correct or remained at a high level of correctness throughout

an exercise day. Using the HDPT visual analytic at the end of exercise days, soldiers correctly predicted a node's ground truth with a 93% accuracy.

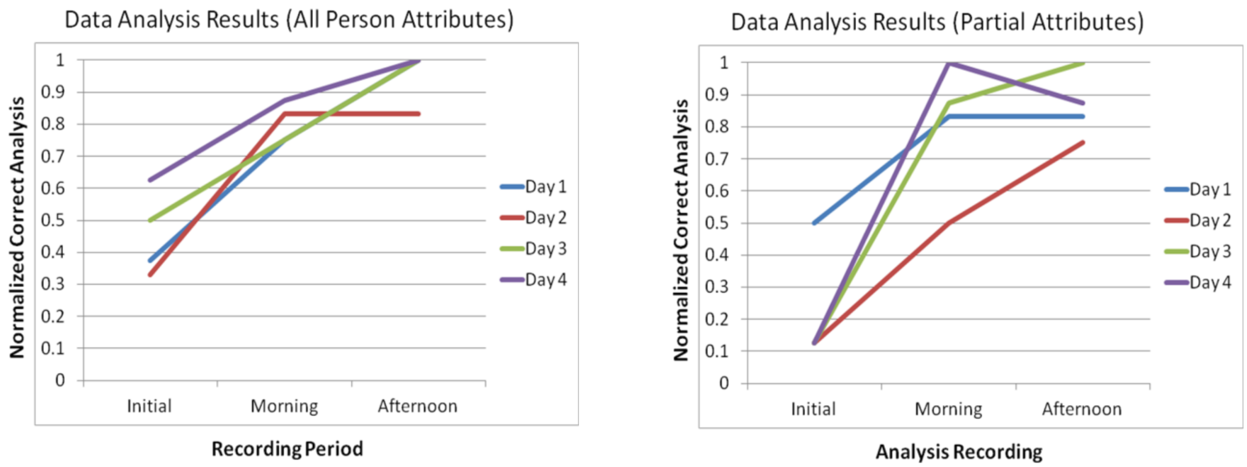


Figure 11: Soldier predictive analysis using the HDPT visual analytic as compared to ground truth.

V. Conclusion

Exploratory data analysis, cluster analysis, pattern recognition, data fusion, data mining …all benefit from visual acuity for the most lucid knowledge extraction from high-dimensional data representations. MDS is a powerful technique holding potential to contribute in all of these areas. Toward that end, HDPT was developed to demonstrate the utility of similarity analysis via a visual analytic in understanding the human terrain. The deployment of HDPT in a tactical environment (Ozone Widget and Global Graph) made it possible to successfully demonstrate this technology and test the power of similarity analysis at the E12 exercise.

An important extension of this work is the development of formal procedures to determine the value of information (VoI) in context of the current operational tempo and the information's content and source reliability [18]. All information should not be weighted equally. Additionally, methods need to be developed for determining appropriate weighting of local attributes and the modeling of linguistic information.

Appendix A: HDPT Survey Form

HDPT C4ISR OTM User Survey
C4ISR MOD On-The-Move Exercise E12

Age: ___ Years MOS: _________ Rank: ________ Time in Service: _____ Yrs

What is your current Duty Position? ____

Are you Color Blind? _____ Computer Use Confidence: 1 2 3 4 5 (1-low; 5-high)

Have you been deployed? _____ if *yes*, number of times _______

Have you received military intelligence training? _____
If *yes*, what courses or informal training?

1. Please circle the number below that best describes how ***well*** you were able to use this tool.

1	2	3	4	5

VERY POOR AVERAGE VERY WELL

2. Please circle the number below that best describes your ***awareness*** of the evolving High Value Individual (HVI) intelligence picture.

1	2	3	4	5

NOT AT ALL AWARE SOMEWHAT AWARE VERY AWARE

3. Please circle the number that best describes the HDPT's ***usefulness*** in tactical deployments.

1	2	3	4	5

NOT AT ALL AWARE SOMEWHAT AWARE VERY AWARE

Comments:

REFERENCES

[1] Thomas, J.J., and Cook K. "Illuminating the Path: The R&D Agenda for Visual Analytics." IEEE Computer Society, p.4. http://nvac.pnl.gov/agenda.stm, 2005

[2] Greitzer, F.L., Noonan, C.F. and Franklin, L.R. 'Cognitive foundations for visual analytics.' Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA, 2011.

[3] Börner, K., Chen, C., & Boyack, K., (2003). Visualizing Knowledge Domains, In Blaise Cronin (Ed.), *Annual Review of Information Science & Technology*, Volume 37,Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, chapter 5, pp. 179-255.

[4] Heer, J., Card, S., & Landay, J. (2005). prefuse: a toolkit for interactive information visualization, Proceedings of the SIGCHI conference on Human factors in computing systems, April 02-07, 2005, Portland, Oregon, USA.

[5] Hanratty, T., Hammell II, R. , Yen, J., McNeese, M., Oh, S., Kim, S., Minotra, D., Strater, L., Cuevas, and Colombo, D. "Knowledge Visualization to Enhance Human-Agent Situation Awareness within a Computational Recognition-Primed Decision System." IEEE Workshop on Situation Management at MILCOM, 2009

[6] Torgerson, W.S. (1952), Multidimensional scaling: I. Theory and method. Psychometrika, 17, 401-419

[7] Young, F., (1985). Understanding Multidimensional Scaling Kotz-Johnson (Ed.) Encyclopedia of Statistical Sciences, Volume 5, Copyright (c) by John Wiley & Sons, Inc accessed online http://forrest.psych.unc.edu/teaching/p208a/mds/mds.html.

[8] Cox, T., & Cox, M. (2001), Multidimensional Scaling, CRC Press 2nd Edition.

[9] Shepard, R. N. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function. Psychometrika, 27, 125-139; 219-246.

[10] Kruskal, J. B. (1964). Nonmetric multidimensional scaling: a numerical method. Psychometrika, 29, 115-129.

[11] Heady, R., & Lucas, J. (2007). PERMAP Operation Manual, University of Louisiana at Lafayette and Agnes Scott College, March 23, 2007.

[12] Stevens, S. S. (1946). "On the Theory of Scales of Measurement". *Science* **103** (2684): 677–680.

[13] Gower, J., (1971). A General Coefficient of Similarity and Some of Its Properties *Biometrics*, Vol. 27, No. 4, pp. 857-871.

[14] Coppock, S. & Mazlack, L. (2004). "Multi-Modal Data Fusion: A Description" to be presented at *KES'2004, 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems* in Wellington, New Zealand.

[15] Ji, Y., Massanari, R., Ager, J., Yen, J., Miller, R., Ying, H. (2007). "A Fuzzy Logic-Based Computational Recognition-Primed Decision Model," *Information Science*.

[16] Podani, J. (1999). Extending Gower's general coefficient of similarity to ordinal characters, Taxon 48, 331-340. May 1999.

[17] Wang, W. (1997). "New similarity measures on fuzzy sets and on elements," *Fuzzy Sets Syst.*, vol. 85, no. 3, pp. 305–309.

[18] Hanratty, T., Hammell II, R., Heilman, E., and Dumer, J., 'Capturing the Value of Information in Complex Military Environments,' IEEE Fuzzy Systems 2012

US ARMY RDECOM

U.S. Army Research, Development and Engineering Command

A Visual Analytic for Improving Human Terrain Understanding

TECHNOLOGY DRIVEN. WARFIGHTER FOCUSED.

Dr. Timothy P. Hanratty, Mr. Eric Heilman, Mr. John Richardson, Mr. Mark Mittrick

18th ICCRTS 19 June 2013

Outline

- Background
- HDPT System Overview
- Scenario for Field Study
- Field Study Method and Results
- Conclusion

Background: Developmental Reasoning

Military decision-making difficulties arise from time critical analysis

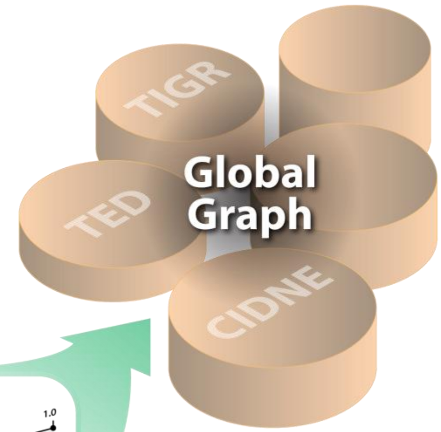
- Of data that
  - Has increased in Amount and types collected
  - Has become inconsistent and incomplete

The Tactical Information and Data Fusion Branch has developed a computerized Tool to

- Assist the military user in quickly analyzing large, inconsistent data sets
- Developed to utilize existing data frameworks

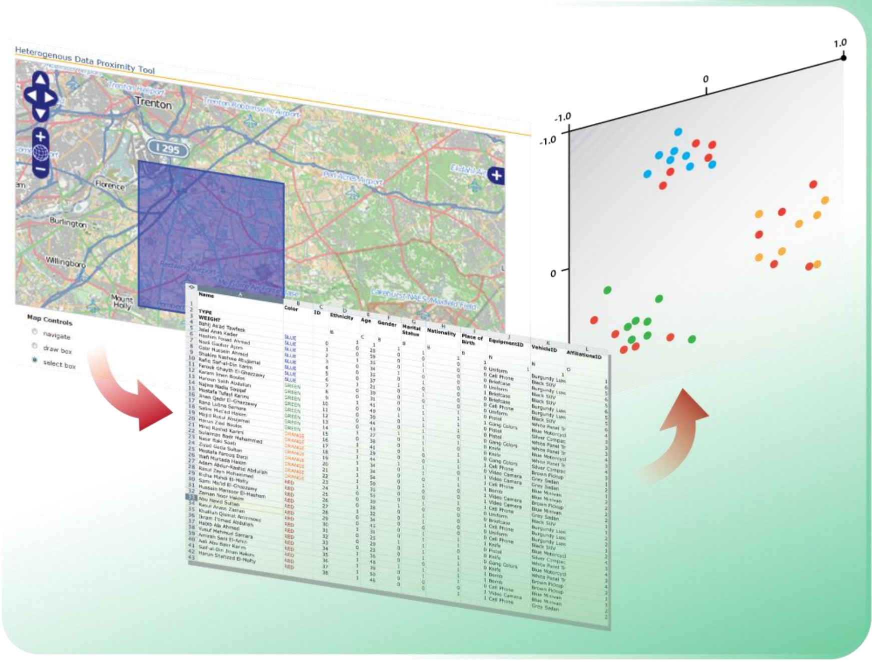
Background: A Solution

The Heterogeneous Data Proximity Tool (HDPT) Uses Multi-Dimensional Scaling (MDS) to reduce high-dimensional data spaces, such human terrain Data, into a human-readable visual analytic



Key Features:

- Tolerant of inconsistent data
- 3D Similarity Display
- Interfaced with DCGS-A OWF
- Mobile Input Device



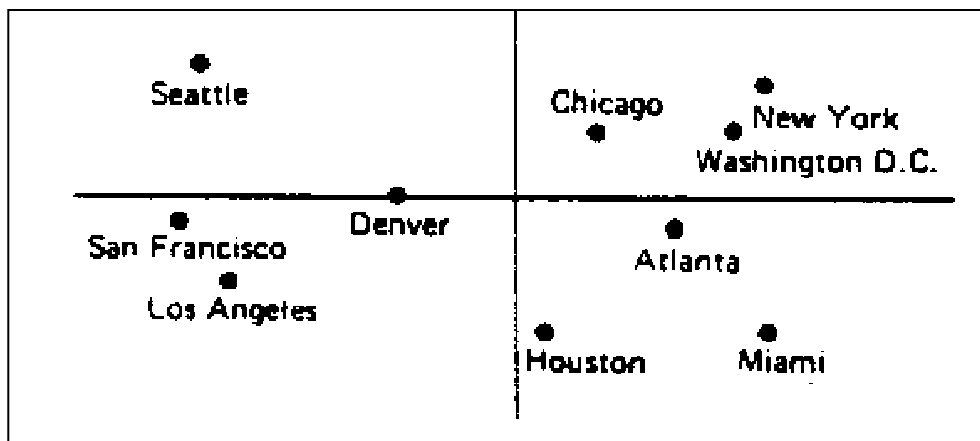
HDPT Algorithms: Multi-Dimensional Scaling

Multidimensional scaling (MDS) is a data analysis approach used to visually interrogate the similarity or dissimilarity between the pair-wise “distances” among a given set of objects.

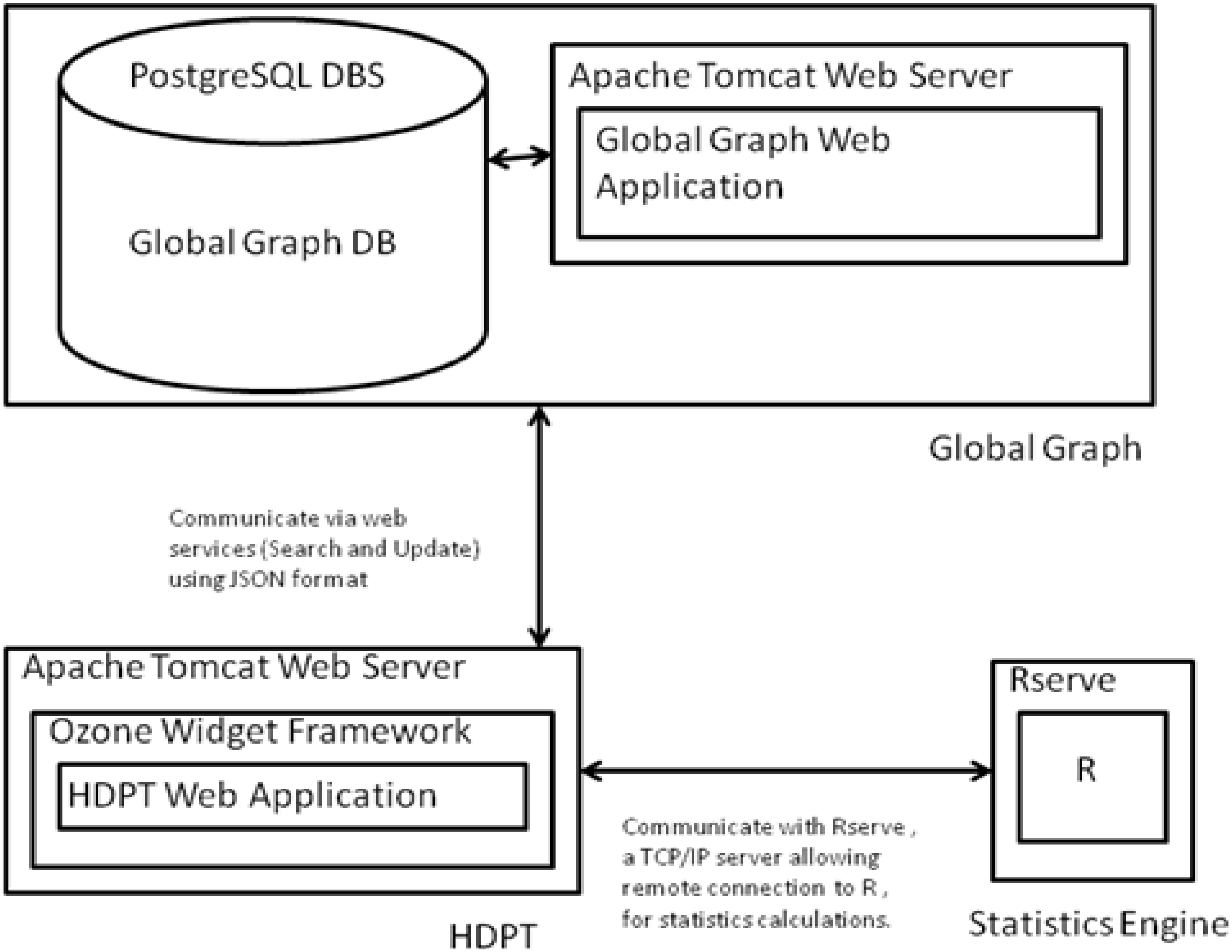
Atlanta	Chicago	Denver	Houston	Los Angeles	Miami	New York	San Francisco	Seattle	Washington, DC	
0	587	1212	701	1936	604	748	2139	2182	543	Atlanta
587	0	920	940	1745	1188	713	1858	1737	597	Chicago
1212	920	0	879	831	1726	1631	949	1021	1494	Denver
701	940	879	0	1374	968	1420	1645	1891	1220	Houston
1936	1745	831	1374	0	2339	2451	347	959	2300	Los Angeles
604	1188	1726	968	2339	0	1092	2594	2734	923	Miami
748	713	1631	1420	2451	1092	0	2571	2408	205	New York
2139	1858	949	1645	347	2594	2571	0	678	2442	San Francisco
2182	1737	1021	1891	959	2734	2408	678	0	2329	Seattle
543	597	1494	1220	2300	923	205	2442	2329	0	Washington, DC

This example from Forrest Young’s *Understanding Multidimensional Scaling* uses the mileage between 10 American cities, shown in the table, as the objective similarity measure.

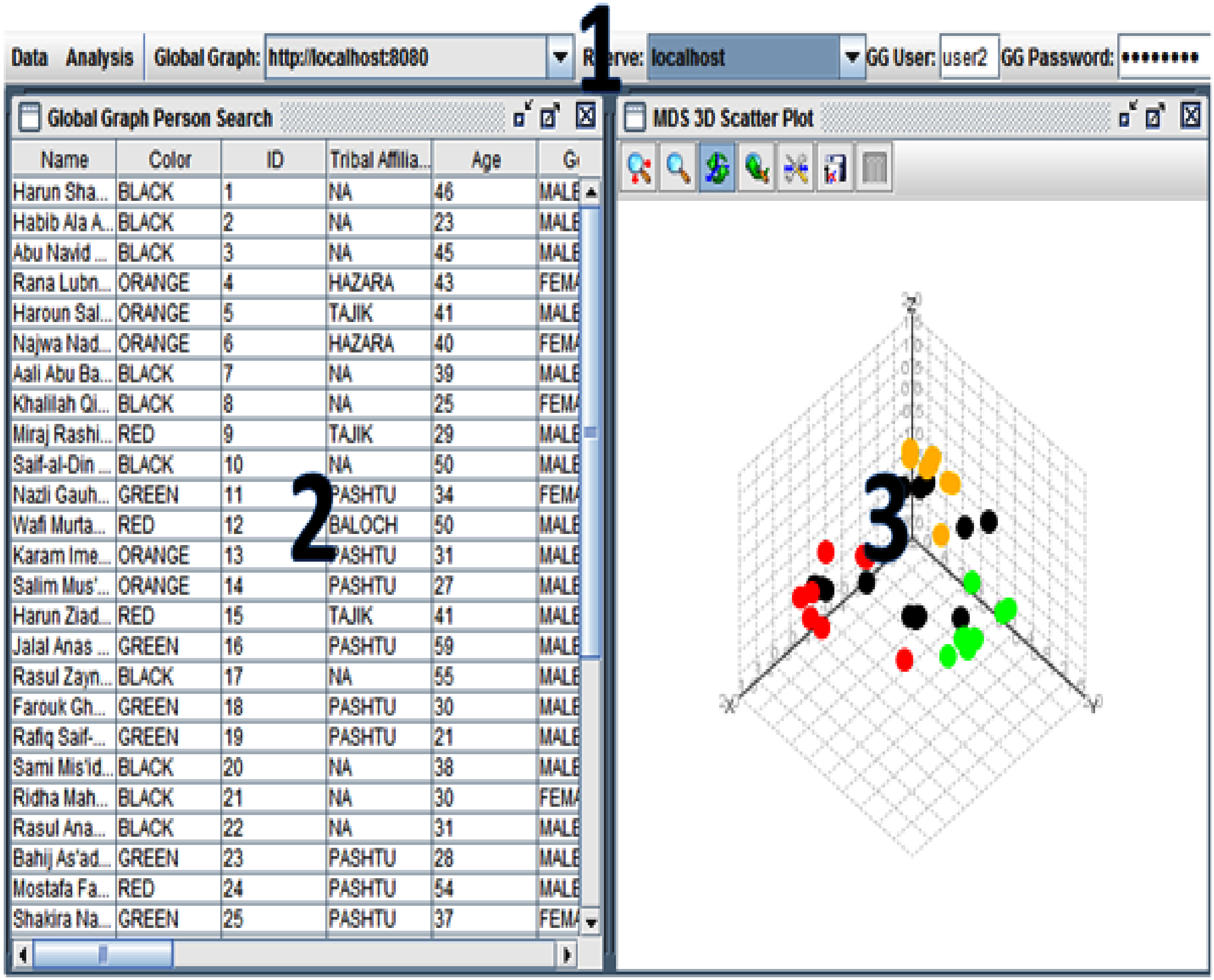
The associated MDS 2-D visualization output would appear something like that found in figure 1. Note: the geometric model allows one to discern the underlying structure and allow human interpretation.



HDPT Component Overview



HDPT Visual Tour



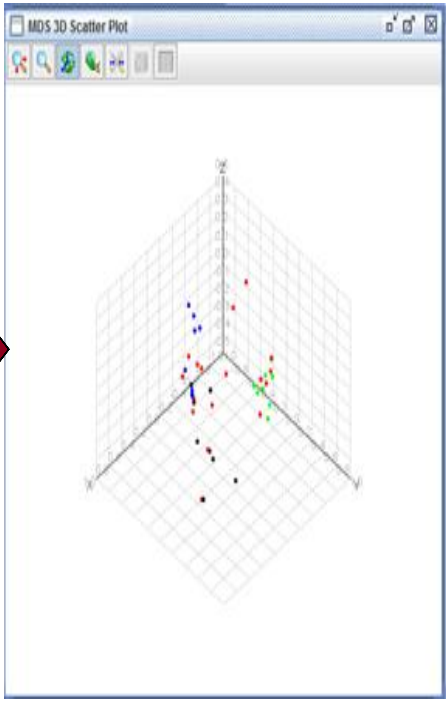
Name	Color	ID	Tribal Affilia...	Age	G
Harun Sha...	BLACK	1	NA	46	MALE
Habib Ala A...	BLACK	2	NA	23	MALE
Abu Navid ...	BLACK	3	NA	45	MALE
Rana Lubn...	ORANGE	4	HAZARA	43	FEM/
Haroun Sal...	ORANGE	5	TAJIK	41	MALE
Najwa Nad...	ORANGE	6	HAZARA	40	FEM/
Aali Abu Ba...	BLACK	7	NA	39	MALE
Khalilah Qi...	BLACK	8	NA	25	FEM/
Miraj Rashi...	RED	9	TAJIK	29	MALE
Saif-al-Din ...	BLACK	10	NA	50	MALE
Nazli Gauh...	GREEN	11	PASHTU	34	FEM/
Wafi Murta...	RED	12	BALOCH	50	MALE
Karam Ime...	ORANGE	13	PASHTU	31	MALE
Salim Mus'...	ORANGE	14	PASHTU	27	MALE
Harun Ziad...	RED	15	TAJIK	41	MALE
Jalal Anas ...	GREEN	16	PASHTU	59	MALE
Rasul Zayn...	BLACK	17	NA	55	MALE
Farouk Gh...	GREEN	18	PASHTU	30	MALE
Rafiq Saif-...	GREEN	19	PASHTU	21	MALE
Sami Mis'id...	BLACK	20	NA	38	MALE
Ridha Mah...	BLACK	21	NA	30	FEM/
Rasul Ana...	BLACK	22	NA	31	MALE
Bahij As'ad...	GREEN	23	PASHTU	28	MALE
Mostafa Fa...	RED	24	PASHTU	54	MALE
Shakira Na...	GREEN	25	PASHTU	37	FEM/

1. **Menu Bar –** Computer communications preferences, data search, and plot controls
2. **Search Window –** Modifiable table of entity attributes
3. **Plot Window –** Rotatable 3D similarity visual analytic

HDPT
C4ISR OTM CONOP

Data describing and individual was collected during morning and afternoon exercise events …

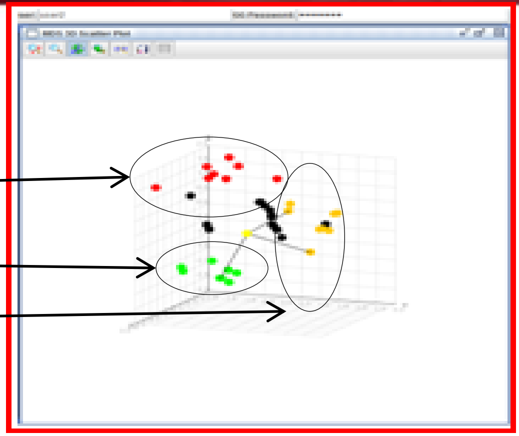
Tribal Affiliation
Marital Status
Nationality
Place of Birth
EquipmentID
VehicleID
Criminal Record
Education Level
Employment Type
Military Record
Religion
Skill
AddressID



Where it was displayed by HDPT for analysis.

And sent to the TOC / DGCS-A Using a mobile device …

HDPT
C4ISR OTM Scenario



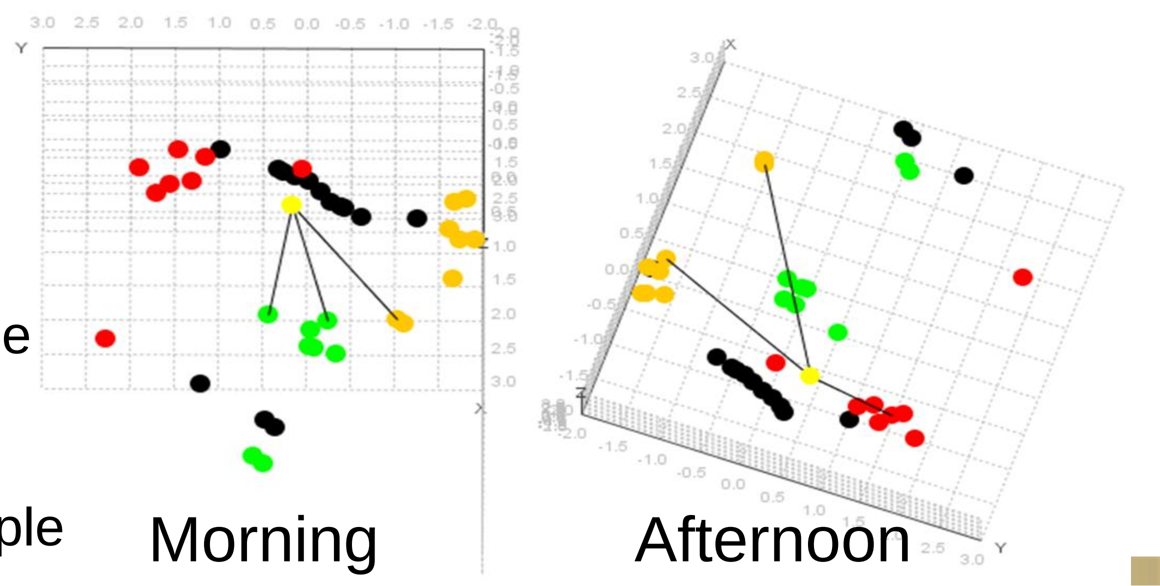
- Each entity is a node in the analytic
- Three reference sets of entities for HDPT:
  - Red or Insurgent
  - Green or Friendly
  - Orange or Criminal
- The black nodes are unknown entities

- Each entity had attributes as shown

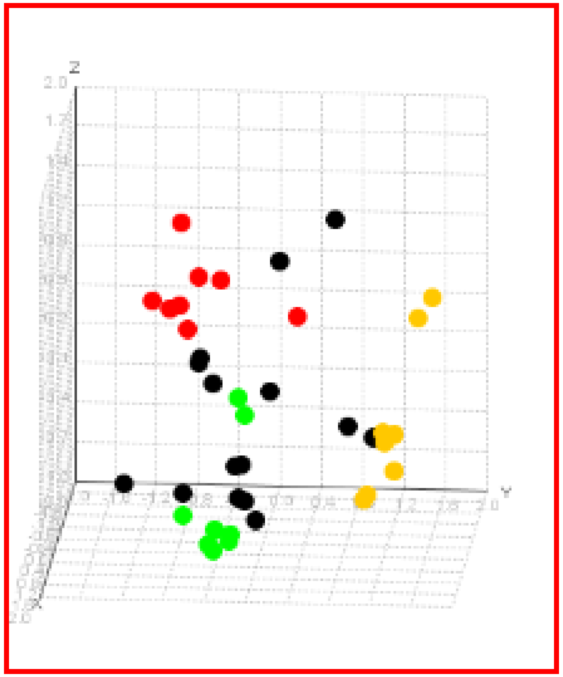
Name	SubID	Location	Org	Age	Gen	TA	MS	Nat	POB	Equip	VehicleID	CR	Ed	Emp	MR	Rel	Sk	Add
Bahij As'ad Tawfeek		39.98 / -74.43	Friendly	28	M	Pastun	M	HN	BIA	Uniform	Burgundy Luxury Sedan	No	High	WC	SS	Mld	ME	TSV

- Unknown entity data was discovered during the exercise

- As attributes were discovered, unknown entities became similar to reference nodes
- Entities could change reference group similarity throughout the exercise day
  - Changing reference example



HDPT Study



Exercise Morning Analytic

From analytics such as that above Soldiers were asked to identify the most likely reference group membership for each unknown entity node

Group Key	
C	Criminal
F	Friendly
I	Insurgent
U	Unknown

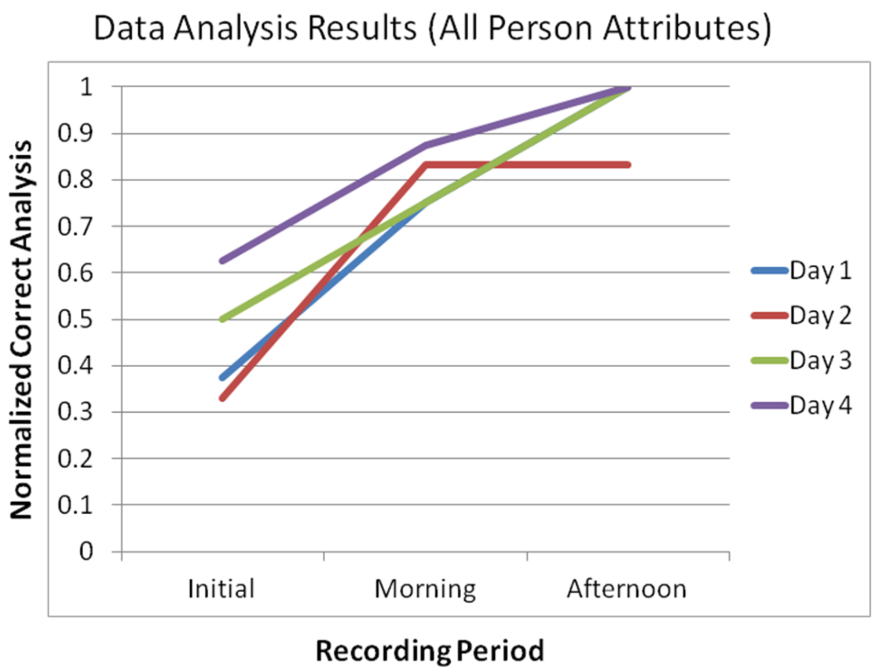
Strength Rating
1 - 5

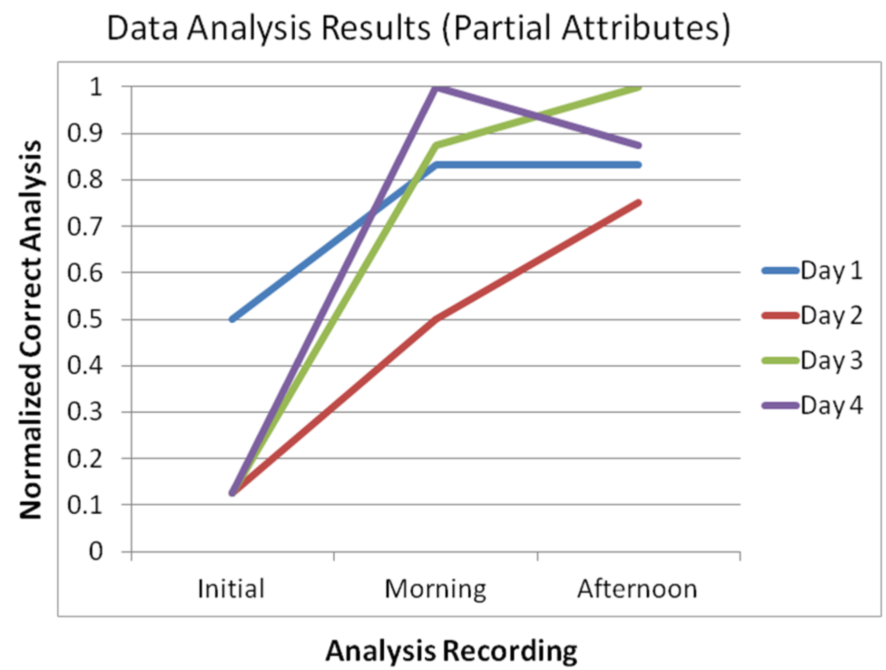
Soldiers rated entity membership using a letter identifying each reference group or a “U” for undetermined and a Likert scale to describe their confidence in the determination

Aali Abu Bakr Karim (S13)	C4
Abu Navid Sultan (S1)	C1
Amirah Sani El-Amin (S12)	U5
Habib Ala Ahmed (S5)	F3
Harun Shahzad El-Mofty (S15)	U5
Hussain Mansoor El-Hashem (S9)	U5
Ikram I'timad Abdullah (S4)	F3
Khalilah Qismat Amirmoez (S3)	U3
Rasul Anass Zaman (S2)	U5
Rasul Zayn Mohammed (S6)	U3
Ridha Mahdi El-Mofty (S7)	F2
Saif-al-Din Jinan Hakim (S14)	I3
Sami Mis'id El-Ghazzawy (S8)	C3
Yusuf Mehmud Samara (S11)	U1
Zaman Noor Hakim (S10)	U5

Completed rating sheets were compared to ground truth entity membership to determine HDPT analysis utility

HDPT Study Results





- Soldier ratings were gathered over 4 exercise days
  - Initialized nodes
  - Morning event nodes
  - Afternoon event nodes
- More attributes were discovered as the exercise day progressed
- Ground truth for unknown nodes was strictly enforced
  - The entire set of attributes was found for some entities (left graph)
  - A partial set of attributes was found for some entities (Right graph)

Soldiers determined correct group affiliation at 93% accuracy.

Conclusion

- HDPT provides a potentially valuable data analysis advantage
  - Soldiers gave high ratings for tactical usefulness and usability
- HDPT provides easy-to-use analytic capabilities for Soldiers untrained in intelligence analysis
- Results support further development and experimentation
  - Plans to return to 2013 C4ISR OTM field exercise with expanded capabilities